# Recent Advances On Ontology Similarity Metrics: A Survey

Georgios Drakopoulos
*Ionian University*
c16drak@ionio.gr

Yorghos Voutos
*Ionian University*
c17vout@ionio.gr

Phivos Mylonas
*Ionian University*
fmylonas@ionio.gr

***Abstract*—With the advent of Semantic Web and the recent advances in the field of knowledge discovery more effort is placed not only on constructing or automatically discovering ontologies but also on determining similarities between ontologies. The latter is usually achieved by a higher order metric taking as input two ontologies in an enhanced graph form and yielding a scalar as a result. This input format allows a considerable degree of flexibility, as the inherently distributed nature of graphs allows the construction of elaborate schemes. Since it is impossible to capture all the essential parameters of an ontology in a single number and given that ontological requirements vary across different domains, it follows that a plethora of such metrics exist. This survey examines the most representative categories of ways for evaluating the similarity between ontology pairs with emphasis placed on the domain of digital culture.**

***Index Terms*—ontologies, semantic graphs, ontology distance, graph mining, digital culture**

## 1. Introduction

Ontologies or knowledge graphs are data structures aiming at capturing in a certain field or application the essence of entities, whether major or minor, and their respective properties and relationships. In a short amount of time ontologies found numerous applications in fields so diverse as the Semantic Web, medical informatics, human genome, political campaigns, user interface and user experience (UI/UX), affective computing, and e-commerce just to name a few. Formally an ontology $J$ is a system represented by the tuple of equation (1):

$$J \triangleq (G, R, A) \tag{1}$$

In the above triplet $G$ is the ground truth set which contains the entities of the underlying domain, $R$ the relationship set which determines the dynamics of that domain, and $A$ is the attribute set which contains all possible relationship types.

Historically the field of ontology dates back as a central branch of ancient Greek philosophy [1] [2] [3] [4] [5] with Parmenides and Aristoteles among others being credited for proposing early ontologies for theology, natural phenomena, and animals in the form of extensive taxonomies. This tradition carried over to modern philosophers such as Martin Heidegger [6], Edmund Husserl [7], and Cornelios Castoriadis [8] who came up with complex concepts and attributes. Now specialized software exists for handling knowledge graphs and efficiently extracting non-trivial conclusions from them. Two significant properties of digital ontologies are [9]:

- **Scale**: Massive knowledge graphs with tens of even hundreds of thousands relationships, predicates, and restrictions between thousands of entities capturing every aspect of a large number of domains or even clusters of related domains are not not uncommon. This is especially true in the case of data-driven ontologies generated by a multitude of sensor readings or other 10V data related applictions.
- **Digitization**: Partly as a sequence of their scale, knowledge graphs need to be understood by computers as well as by humans of various capacities like fact checkers, data analysts, opinion leaders, and domain experts. To this end, a plethora of representations has beeen developed for the former case and various visualization techniques for the latter.

Given the above properties, it is consequently only logical to ask whether it would be possible to compare any two ontologies. This question arises from the following three general reasons, without considering additional application-specific motivations for performing a comparison [10]:

- An appropriately designed metric can reveal how well a given ontology captures the essence of the underlying field. This is essential since different knowledge graphs may well model the same domain in a different but not necessarily equivalent way.
- Instead of directly clustering objects coming from a given domain, ontologies for it can be clustered first and then objects can in turn be clustered according to how well they fit to these ontologies, perhaps in a fuzzy way for additonal flexibility.
- Comparisons between compatible ontologies coming from related or in certain cases even remotely connected fields may allow for knowledge transfer across domains. This strategy capitalizes on already established knowledge for paving the way for breakthroughs in other domains than the original one.

Digital cultural and cultural preservation are prime application fields for ontologies since, like almost any human activity, they are highly structured with a plethora of latent

978-1-7281-6445-8/20/$31.00 ©2020 IEEE

higher order patterns [11]. With the recent sharp rise of the need for high quality cultural products, ontologies tailored for these domains can serve as the principal building blocks for successful and scalable cultural analytics such as those presented among others in [12], [13], [14], [15], and [16].

The primary research contribution of this conference paper is twofold. First, a critical overview of the distance metrics designed for ontologies found in the recent scientific literature is given. From this a set of requirements for knowledge graph similarity metrics as well as general guidelines for selecting such metrics in various engineering scenarios are given. Second, a considerable number of literature references about ontology applications and ontology alignment is given so that this conference paper can serve as a starting point for any research in the field.

The reamaining of this work is structured as follows. In section 2 the respective bodies of work regarding the topics of graph mining, Semantic Web, graph signal processing, persistent data structures, and social network analysis are summarized. Ontological applications are covered in detail in section 3. The specifics of ontology metrics is described in section 4 and recommendations based on it for certain significant engineering applications are given in 5. The main points of this work are summarized in section 6. Technical acronyms are explained the first time they are encountered in the text. Finally, the notation of this conference paper is summarized in table 1.

TABLE 1. Notation of this conference paper.

| Symbol | Meaning |
|---|---|
| $\triangleq$ | Definition or equality by definition |
| $\{s_1, \ldots, s_n\}$ | Set containing elements $s_1, \ldots, s_n$ |
| $(t_1, \ldots, t_n)$ | Tuple with elements $t_1, \ldots, t_n$ |
| $/s/$ | String $s$ |
| $\lvert S \rvert$ | Set or tuple cardinality or string length |

## 2. Previous Work

The field of graph mining frequently provides the algorithmic tools for representing and clustering ontologies represented as graphs [17]. Community structure discovery or graph partitioning is a significant problem with a plethora of applications [18]. The discovery process may well rely on local connectivity patterns such as degrees, triangles, and squares as shown in [19] and in [20]. Kernel approaches [21] are also viable approaches to graph clustering. More recently, increased computational power has allowed the use of local higher order connectivity properties as explained in [22]. Graph spectral properties also rely heavily on structural patterns but in a linear algebraic way [23] [24]. Alternatively, graph partitioning can be accomplished with functional properties such as flow simulation [25], cross-correlation of functional attributes [26], or multiscale decomposition of graph functions to wavelets [27]. These properties may be better suited for a given scenario [28], but structural ones can be applied to virutally any graph. A summary of the various methodologies is given in [29].

Recently, graph signal processing has stimulated research interest in ontologies [30]. Instead of considering graphs as combinatorial objects, they are treated as two-dimensional deterministic or stochastic signals [31]. Prime examples include graph sampling [32] and graph Laplacian inverse estimation [33]. A probabilistic framework for graph signal processing is described in [34]. In [35] a tensor stack network (TSN) is trained to evaluate the structural resilience of communication networks. This approach stems from the linear algebraic representation of graphs through their respective adjacency matrices [36]. The latter can reveal properties such as higher order connectivity patterns [37] and vertex centrality [38]. Alternatively, the graph Laplacian [39] or quasi-Laplacian [40] can be used to discover similar properties. Multilayer or multiplex graphs allow the existence of multiple edges between the same vertices as long as the edges have distinct labels [41]. Spectral clustering with convex optimization for multilayer graphs is explored in [42]. Graph signal processing applications include brain functional imaging [43] and reconstructing network topology from sampled data packet activities [44].

A significant part of the Semantic Web is also built on ontologies [45]. In [46] various Semantic Web ontologies are categorized in terms of semantic complexity. A review of programming languages for the Semantic Web is given in [47] and also in [48]. Graph theoretic databases [49] are an integral part of the NoSQL movement along with document databases such as MongoDB [50], column family stores like Cassandra [51] [52], and key-value or associative array databases such as Redis [53]. They are also employed in Semantic Web applications as shown in the case of Neo4j [54] and TitanDB [55]. Fuzzy graph queries extend the capabilities of these ontologies [56]. Moreover, the recent advent of large distributed processing tools such as Apache Spark [57], which relies on the Hadoop Distributed File System (HDFS) [58], resulted in a paradigm shift. For instance the MLlib library [59] of Apache Spark contains numerous graph processing algorithms [60]. An overview of recent systems and frameworks for massive graph processing can be found among others in [61].

Persistent data structures [62] have the remarkable property that they can provide efficient access to their past versions [63]. Amortization for persistent lists is explored in [64]. A space efficient and persistent data structure for representing property graphs allowing rollback in graph databases is described in [65]. Declarative languages like Haskell [66] are ideal for discovering recursive patterns in graphs as well as for executing concurrent operations on them [67]. Hardware design with Haskell is explored in [68], whereas computation acceleration with multiple graphics processing units (GPUs) is explored in [69] and in [70].

Social network analysis also relies heavily on graph mining [71]. Strategies for heterogeneous social networks are explored in [72]. The detection of polarity shifts in Twitter conversations through the Hilbert-Huang spectrum of the affective contents of the tweets in these conversation is proposed in [73]. Automated ontology discovery in social networks through machine learning (ML) is described in

[74]. Natural language processing (NLP) techniques play a central role in discovering entities, attributes, and relationships in ontologies [75].

# 3. Applications

Over a short time span an impressive number of knowledge graphs has been developed for a broad array of applications. This has led not only to a deeper understanding of them, but also in certain cases in non-trivial knowledge transfer between them. In turn, this resulted in reinforced interdisciplinary ties.

Genes because of their structure, variations in functionality, and importance have been modeled multiple times as in [76] where gene functionality is tied to their geometrical properties expressed as manifolds. In [77] gene distances based on functional attributes such as the number of aminoacids are proposed. Moreover, another family of gene similarity defined over the cross-correlation of gene expressions is described in [78].

In computer engineering an ontology expressed in resource description framework (RDF) for describing computer networks is described in [79]. Additionally, similarity metrics for ontologies designed to represent Semantic Web services are described in [80]. Text clustering assisted by ontologies for text semantics is proposed in [81].

The domains of digital culture and cultural heritage are paramount in the digital world of today. The increasing demand for high quality cultural products is a major driver behind this trend. A JSON ontology for the cultural heritage of the Greek region of Ionian Islands taking into account aspects such as artistic trends and linguistic properties is presented in [82]. A scheme for the effective mining cultural attributes from text is proposed in [83]. The blueprints of an ontology based system for matching the needs of tourists utilizing group packages are given in [84]. Finally, an ontology for Chinese advertisements in terms of cultural references is described in [85].

# 4. Distance Metrics

## 4.1. Alignment, Representations, And Distances

Aligning two or more ontologies is the process of matching, perhaps with some uncertainty, their respective entities, relationships, and attributes. It is a necessary step before the application of any ontology similarity metric if meaningful results are to be obtained. Due to its importance, many alignment schemes can be found. It is worth mentioning that their vast majority depends on the chosen representation. The most common of them in abstact form are the RDF triplets [86], graphs [87], and strings [88]. Moreover, standard data description formats such as those shown in table 2 can be adapted to represent ontologies.

TABLE 2. DATA DESCRIPTION FORMATS.

| Name | Data type |
|---|---|
| JSON | Description schema for generic data |
| JSON-LD | Description for graphs and linked data |
| BSON | Binary object descripion scehema |
| XML | Structured and semantic data schema |

From a graph theory perspective, ontology alignment is the art of finding a suitable homomorphism such that vertices representing the same entities are matched. In this case, assuming that the entities and the relationships of the two knowledge graphs are contained respectively in the pair $G_1$ and $G_2$ and in the pair $R_1$ and $R_2$, then typically a cost function of the following form is mimimized:

$$K \triangleq g(G_1, G_2) + \lambda_1 g'(R_1, R_2) + \lambda_2 |G_1 \setminus G_2| + \lambda_3 |R_1 \setminus R_2| \quad (2)$$

In equation (2) the hyperparameters $\lambda_1$, $\lambda_2$, and $\lambda_3$ indicate the relative weight with respect to the first term. The last two terms penalize longer matchings in case there is no exact correspondence between the two pairs of sets in the same way the penalty terms in metrics like the Aikake Information Criterion (AIC) or the Bayes Infromation Criterion (BIC) exclude longer probabilistic explanations of a given set of observations. Moreover, $g(\cdot, \cdot)$ and $g(\cdot, \cdot)'$ are distance metrics for each set pair.

When a string representation is selected, then an alphabet matching is performed. In this case, both ontologies are encoded as the strings $/w_1/$ and $/w_2/$ and the cost function may take the form:

$$Q \triangleq h(/w_1/\,, /w_2/) + \mu_1 |/w_1/ - /w_2/| \quad (3)$$

In equation (3) $\mu_1$ is a hyperparameter serving the same purpose as before and $h(\cdot, \cdot)$ is a string similarity metric.

The above can cen be achieved with strategies such as these proposed in [89], [90], [91], or [92]. Ontology alignment eliminates or mitigates the following challenges:

- In case there is ambiguity over entity or relationship matching, then it is minimized according to a prespecified criterion which may well include a small number of hyperparameters.
- When there are multiple equivalent representations for the same domain, then the shortest one is chosen.
- Missing relationships or entities are penalized.

Maintaining connections for linked data is described in [93]. Storing RDF triplets to NoSQL databses is considered in [94]. Uncertain reasoning with these triplets is examined in [95]. String metrics for ontology comparisons are studied in [96]. Along a similar line of reasoning with string operations moves [97]. In [98] the nature of linked data is taken into consideration in order to yield an alignment scheme.

Precision and recall in a semantic context are defined in [99] for ontology alignment. Correspondence patterns are extracted with ML models and the resulting knowledge graphs are studied in [100]. Discrete particle swarm optimization for massive ontologies alignment is described in [101]. Concerning cultural heritage ontologies, the alignment of RDF triplets describing cultural data is discussed in [102]. User validation for ontology alignment is explored in [103].

Because of the intense interdisciplinary interest in ontologies, a number of taxonomies have been already proposed in the relevant scientific literature like these in [104] and in [105]. Semantically enriched distances between pairs of knowledge graphs are presented in [106] as well as in [107]. Learning strategies for training ML models to gradually learn personal ontologies are descibed in [108], while ML models for hierarchical ontologies are proposed in [109]. A matrix learning approach is given in detail in [110]. A thorough and overall review of ontology distances across various domains is given in [111]. Also, many of the above strategies are collected and described in detail among others in [112] and in [113].

### 4.2. Metric Requirements

Given that ontologies are increasingly becoming more detailed and complex, primarily in terms of relationships and attributes, it makes sense to find distance metrics which not only are scalable but they also discover efficiently higher order patterns. In brief, any procedure for constructing meaningful massive knowledge graphs out of domains with possible uncertainty, should meet at least the following list of requirements:

- When there are multiple or composite types of relationships between the ground truth entities or when the ontology is gradually constructed in a data-driven context, then an adaptive algorithmic scheme which takes into account a recent window of representation errors such as the one proposed in [114] should be used. Although modeling is done locally, in the long term it is an efficient process when large data volumes are involved.
- For alternative representations the distance between them should be bounded, meaning that the order of data arrival should play little role in the eventual ontology formulation.
- Concerning the case where uncertainty is involved in the relationships, it should be coded as probabilities which are part of the resulting ontology.

## 5. Recommendations

The analysis of the preceding sections lead to certain recommendations for selecting ontology distance metrics. It should come as no surprise that in the emerging era of 10V data many ontologies are constructed in a data-driven manner from abstractions generated from massive information content coming from multiple sources such as internet of things (IoT) micronetworks, finance and business [115], smart city sensor arrays, computational biology models [116], or digital health mobile applications [117]. As a general rule, corroborated in part by the arguments found in the scientific literature, the following should hold:

- The efficiency-accuracy tradeoff should be either favor the former or, even better, be dynamic depending on local data properties.
- Emphasis should be placed on robustness in the sense that frequently arising patterns should be taken more into consideration.

## 6. Conclusions And Future Work

This conference paper focuses on similarity metrics for ontologies with an emphasis on the cultural heritage and digital culture domains. Specifically, it sets forth a set of requirements for the class of metrics taking pairs knowledge graphs and mapping them to a single scalar. Moreover, based on the extensive review of the recent scientific literature regarding ontology distance metrics a set of practical recommendations for various engineering scenarios in the emerging era of 10V data is given.

Concerning future work directions in the field of knowledge graphs the following can be said. Computing aspects should include the development of space efficient representations for massive ontologies with a large number of sparse relationships, perhaphs based on existing know-how of sparse tensor representations. Moreover, the full or partial discovery of missing attributes, entities, or relationships with link prediction or estimation theory techniques should be investigated. Alternatively, crowdsourcing techniques or information extracted from social networks can be used for that purpose. Since ontologies intersect even in part with various sectors of human activity, a third line of research should be the development of sophisticated systems for facilitating the input of mutiple domain experts, perhaps with visualization of complex notions. As human input can very well be fuzzy or contain contradictions, various robust algorithmic techniques for discovering and resolving these should be also pursued.

## Acknowledgment

This research has been co-financed by the European Union and Greek national funds through the Competitiveness, Entrepreneurship, and Innovation Operation Programme under the call “Research – Create – Innovate” project title “Development of technologies and methods for cultural inventory data interoperability”, project code T1EDK01728, MIS code 5030954.

## References

[1] G. E. Lloyd, “Humanity between gods and beasts? Ontologies in question,” *Journal of the Royal Anthropological Institute*, vol. 17, no. 4, pp. 829–845, 2011.

[2] D. Van der Schyff, “On being and becoming: Ancient Greek ethics and ontology in the twenty-first century,” Ph.D. dissertation, SFU, 2010.

[3] C. Poster, “Being and becoming: Rhetorical ontology in early Greek thought,” *Philosophy & Rhetoric*, pp. 1–14, 1996.

[4] W. Brogan, "Double archê: Heidegger's reading of Aristotle's kinetic ontology," *Angelaki: Journal of the theoretical humanities*, vol. 11, no. 3, pp. 85–92, 2006.

[5] J. R. Searle, "Language and social ontology," *Theory and Society*, vol. 37, no. 5, pp. 443–459, 2008.

[6] J. Paley, "Misinterpretive phenomenology: Heidegger, ontology and nursing research," *Journal of advanced nursing*, vol. 27, no. 4, pp. 817–824, 1998.

[7] D. Zahavi, *Husserl's phenomenology*. Stanford University Press, 2003.

[8] S. Adams, *Castoriadis's ontology: Being and creation*. Fordham Univ Press, 2011.

[9] F. Kharbat and H. El-Ghalayini, *Building ontology from knowledge base systems*. INTECH Open Access Publisher, 2008.

[10] L. U. Ulivi, "First steps toward a systemic ontology," in *Systemics of Incompleteness and Quasi-Systems*. Springer, 2019, pp. 57–73.

[11] M. Deuze, "Participation, remediation, bricolage: Considering principal components of a digital culture," *The information society*, vol. 22, no. 2, pp. 63–75, 2006.

[12] L. Manovich, "Cultural analytics: Visualising cultural patterns in the era of more media," *Domus March*, 2009.

[13] ——, "How to follow global digital cultures, or cultural analytics for beginners," *Deep Search: The politics of search beyond Google*, 2009.

[14] S. Yamaoka, L. Manovich, J. Douglass, and F. Kuester, "Cultural analytics in large-scale visualization environments," *Computer*, vol. 44, no. 12, pp. 39–48, 2011.

[15] G. Drakopoulos, I. Giannoukou, P. Mylonas, and S. Sioutas, "The converging triangle of cultural content, cognitive science, and behavioral economics," in *MHDW*, vol. 585. Springer, 2020, pp. 200–212.

[16] N. G. Smith, S. Cutchin, R. Kooima, R. A. Ainsworth, D. J. Sandin, J. Schulze, A. Prudhomme, F. Kuester, T. E. Levy, and T. A. DeFanti, "Cultural heritage omni-stereo panoramas for immersive cultural analytics – From the Nile to the Hijaz," in *ISPA*. IEEE, 2013, pp. 552–557.

[17] S. E. Schaeffer, "Graph clustering," *Computer science review*, vol. 1, no. 1, pp. 27–64, 2007.

[18] Y. Zhou, H. Cheng, and J. X. Yu, "Graph clustering based on structural/attribute similarities," *PVLDB*, vol. 2, no. 1, pp. 718–729, 2009.

[19] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[20] J. Yang and J. Leskovec, "Community-affiliation graph model for overlapping network community detection," in *ICDM*. IEEE, 2012, pp. 1170–1175.

[21] B. Kulis, S. Basu, I. Dhillon, and R. Mooney, "Semi-supervised graph clustering: A kernel approach," *Machine Learning*, vol. 74, no. 1, pp. 1–22, 2009.

[22] H. Yin, A. R. Benson, J. Leskovec, and D. F. Gleich, "Local higher-order graph clustering," in *KDD*, 2017, pp. 555–564.

[23] I. S. Dhillon, "Co-clustering documents and words using bipartite spectral graph partitioning," in *KDD*, 2001, pp. 269–274.

[24] C. Dhanjal, R. Gaudel, and S. Clémençon, "Efficient eigen-updating for spectral graph clustering," *Neurocomputing*, vol. 131, pp. 440–452, 2014.

[25] S. M. Van Dongen, "Graph clustering by flow simulation," Ph.D. dissertation, Center for math and computer science (CWI), 2000.

[26] S. Dodel, J. M. Herrmann, and T. Geisel, "Functional connectivity by cross-correlation clustering," *Neurocomputing*, vol. 44, pp. 1065–1070, 2002.

[27] A. Antoniadis, X. Brossat, J. Cugliari, and J.-M. Poggi, "Clustering functional data using wavelets," *International Journal of Wavelets, Multiresolution, and Information Processing*, vol. 11, no. 01, 2013.

[28] D. Kuang, C. Ding, and H. Park, "Symmetric nonnegative matrix factorization for graph clustering," in *SDM*. SIAM, 2012, pp. 106–117.

[29] U. Brandes, M. Gaertler, and D. Wagner, "Experiments on graph clustering algorithms," in *European Symposium on Algorithms*. Springer, 2003, pp. 568–579.

[30] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *ICASSP*. IEEE, 2013, pp. 5445–5449.

[31] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[32] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE transactions on signal processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[33] E. Pavez and A. Ortega, "Generalized Laplacian precision matrix estimation for graph signal processing," in *ICASSP*. IEEE, 2016, pp. 6350–6354.

[34] C. Zhang, D. Florêncio, and P. A. Chou, "Graph signal processing - A probabilistic framework," Microsoft Research, Redmond, WA, Tech. Rep. MSR-TR-2015-31, 2015.

[35] G. Drakopoulos and P. Mylonas, "Evaluating graph resilience with tensor stack networks: A Keras implementation," *NCAA*, pp. 1–16, 2020.

[36] R. L. Rivest and J. Vuillemin, "On recognizing graph properties from adjacency matrices," *Theoretical Computer Science*, vol. 3, no. 3, pp. 371–384, 1976.

[37] E. Estrada and D. J. Higham, "Network properties revealed through matrix functions," *SIAM Review*, vol. 52, no. 4, pp. 696–714, 2010.

[38] E. Nathan and D. A. Bader, "Incrementally updating Katz centrality in dynamic graphs," *SNAM*, vol. 8, no. 1, 2018.

[39] X. Qi, E. Fuller, Q. Wu, Y. Wu, and C.-Q. Zhang, "Laplacian centrality: A new centrality measure for weighted networks," *Information Sciences*, vol. 194, pp. 240–253, 2012.

[40] Y. Ma, Z. Cao, and X. Qi, "Quasi-Laplacian centrality: A new vertex centrality measurement based on Quasi-Laplacian energy of networks," *Physica A: Statistical mechanics and its applications*, vol. 527, 2019.

[41] Y. Yang, S. Han, T. Wang, W. Tao, and X.-C. Tai, "Multilayer graph cuts based unsupervised color–texture image segmentation using multivariate mixed Student's t-distribution and regional credibility merging," *Pattern Recognition*, vol. 46, no. 4, 2013.

[42] P.-Y. Chen and A. O. Hero, "Multilayer spectral graph clustering via convex layer aggregation: Theory and algorithms," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 553–567, 2017.

[43] W. Huang, T. A. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 868–885, 2018.

[44] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.

[45] A. Gómez-Pérez and O. Corcho, "Ontology languages for the Semantic Web," *IEEE Intelligent Systems*, vol. 17, no. 1, pp. 54–60, 2002.

[46] J. Seidenberg and A. Rector, "Web ontology segmentation: Analysis, classification and use," in *5th International conference on World Wide Web*, 2006, pp. 13–22.

[47] G. Antoniou, E. Franconi, and F. Van Harmelen, “Introduction to Semantic Web ontology languages,” in *Reasoning Web*. Springer, 2005, pp. 1–21.

[48] J. Pulido, M. Ruiz, R. Herrera, E. Cabello, S. Legrand, and D. Elliman, “Ontology languages for the Semantic Web: A never completely updated review,” *Knowledge-Based Systems*, vol. 19, no. 7, pp. 489–497, 2006.

[49] I. Robinson, J. Webber, and E. Eifrem, *Graph databases: New opportunities for connected data*. O’Reilly Media Inc., 2015.

[50] K. Banker, *MongoDB in action*. Manning Publications Co., 2011.

[51] A. Lakshman and P. Malik, “Cassandra: Structured storage system on a p2p network,” in *PODC*, 2009.

[52] ——, “Cassandra: A decentralized structured storage system,” *ACM SIGOPS Operating Systems Review*, vol. 44, no. 2, pp. 35–40, 2010.

[53] T. Macedo and F. Oliveira, *Redis cookbook: Practical techniques for fast data manipulation*. O’Reilly Media Inc, 2011.

[54] J. Webber, “A programmatic introduction to Neo4j,” in *3rd annual conference on systems, programming, and applications: Software for humanity*, 2012, pp. 217–218.

[55] G. Drakopoulos, A. Kanavos, P. Mylonas, S. Sioutas, and D. Tsolis, “Towards a framework for tensor ontologies over Neo4j: Representations and operations,” in *IISA*. IEEE, 2017.

[56] O. Pivert, V. Thion, H. Jaudoin, and G. Smits, “On a fuzzy algebra for querying graph databases,” in *ICTAI*. IEEE, 2014, pp. 748–755.

[57] M. Zaharia *et al.*, “Apache Spark: A unified engine for big data processing,” *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.

[58] J. Shafer, S. Rixner, and A. L. Cox, “The Hadoop distributed filesystem: Balancing portability and performance,” in *ISPASS*. IEEE, 2010, pp. 122–133.

[59] X. Meng *et al.*, “MLlib: Machine learning in Apache Spark,” *JMLR*, vol. 17, no. 1, pp. 1235–1241, 2016.

[60] B. Quinto, “Introduction to Spark and Spark MLlib,” in *Next-Generation Machine Learning with Spark*. Springer, 2020, pp. 29–96.

[61] S. Aridhi and E. M. Nguifo, “Big graph mining: Frameworks and techniques,” *Big Data Research*, vol. 6, pp. 1–10, 2016.

[62] J. R. Driscoll, N. Sarnak, D. D. Sleator, and R. E. Tarjan, “Making data structures persistent,” *Journal of computer and system sciences*, vol. 38, no. 1, pp. 86–124, 1989.

[63] A. Dearle, Q. Cutts, and G. Kirby, “Browsing, grazing and nibbling persistent data structures,” in *Persistent Object Systems*. Springer, 1990, pp. 56–69.

[64] C. Okasaki, “Amortization, lazy evaluation, and persistence: Lists with catenation via lazy linking,” in *36th Annual Foundations of Computer Science*. IEEE, 1995, pp. 646–654.

[65] S. Kontopoulos and G. Drakopoulos, “A space efficient scheme for persistent graph representation,” in *ICTAI*. IEEE, 2014, pp. 299–303.

[66] P. Hudak and J. H. Fasel, “A gentle introduction to Haskell,” *ACM SIGPLAN Notices*, vol. 27, no. 5, pp. 1–52, 1992.

[67] S. Thompson, *Haskell: The craft of functional programming*. Addison-Wesley, 2011, vol. 2.

[68] P. Bjesse, K. Claessen, M. Sheeran, and S. Singh, “Lava: Hardware design in Haskell,” *ACM SIGPLAN Notices*, vol. 34, no. 1, pp. 174–184, 1998.

[69] T. Sheard and S. P. Jones, “Template meta-programming for Haskell,” in *ACM SIGPLAN workshop on Haskell*, 2002, pp. 1–16.

[70] M. M. Chakravarty, G. Keller, S. Lee, T. L. McDonell, and V. Grover, “Accelerating Haskell array codes with multicore GPUs,” in *Sixth workshop on declarative aspects of multicore programming*, 2011, pp. 3–14.

[71] D. Knoke and S. Yang, *Social network analysis*. Sage Publications, 2019, vol. 154.

[72] D. Cai, Z. Shao, X. He, X. Yan, and J. Han, “Mining hidden community in heterogeneous social networks,” in *3rd international workshop on link discovery*, 2005, pp. 58–65.

[73] G. Drakopoulos, A. Kanavos, P. Mylonas, and S. Sioutas, “Discovering sentiment potential in Twitter conversations with Hilbert–Huang spectrum,” *EVOS*, pp. 1–15, 2020.

[74] A. B. F. Mansur and N. Yusof, “Social learning network analysis model to identify learning patterns using ontology clustering techniques and meaningful learning,” *Computers & Education*, vol. 63, pp. 73–86, 2013.

[75] H. Chen and X. Luo, “An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing,” *Advanced Engineering Informatics*, vol. 42, 2019.

[76] G. Lerman and B. E. Shakhnovich, “Defining functional distance using manifold embeddings of gene ontology annotations,” *PNAS*, vol. 104, no. 27, pp. 11 334–11 339, 2007.

[77] A. Del Pozo, F. Pazos, and A. Valencia, “Defining functional distances over gene ontology,” *BMC Bioinformatics*, vol. 9, no. 1, p. 50, 2008.

[78] H. Wang, F. Azuaje, O. Bodenreider, and J. Dopazo, “Gene expression correlation and gene ontology-based similarity: An assessment of quantitative relationships,” in *Symposium on Computational Intelligence in Bioinformatics and Computational Biology*. IEEE, 2004, pp. 25–31.

[79] J. J. Van der Ham, F. Dijkstra, F. Travostino, H. M. Andree, and C. T. de Laat, “Using RDF to describe networks,” *Future Generation Computer Systems*, vol. 22, no. 8, pp. 862–867, 2006.

[80] X. Wang, Y. Ding, and Y. Zhao, “Similarity measurement about ontology-based Semantic Web services,” in *SemWS*, 2006.

[81] L. Jing, L. Zhou, M. K. Ng, and J. Z. Huang, “Ontology-based distance measure for text clustering,” in *SIAM SDM workshop on text mining*, 2006.

[82] G. Drakopoulos, E. Spyrou, Y. Voutos, and P. Mylonas, “A semantically annotated JSON metadata structure for open linked cultural data in Neo4j,” in *PCI*. ACM, 2019, pp. 81–88.

[83] M. Whitelaw, M. Guglielmetti, and T. Innocent, “Strange ontologies in digital culture,” *CIE*, vol. 7, no. 1, pp. 1–13, 2009.

[84] D. N. Kanellopoulos, “An ontology-based system for intelligent matching of travellers’ needs for group package tours,” *International Journal of Digital Culture and Electronic Tourism*, vol. 1, no. 1, pp. 76–99, 2008.

[85] H.-T. Zheng, J.-Y. Chen, and Y. Jiang, “An ontology-based approach to Chinese semantic advertising,” *Information Sciences*, vol. 216, pp. 138–154, 2012.

[86] M. Klein, “Interpreting XML documents via an RDF schema ontology,” in *International Workshop on Database and Expert Systems Applications*. IEEE, 2002, pp. 889–893.

[87] S. Pouriyeh, M. Allahyari, Q. Liu, G. Cheng, H. R. Arabnia, M. Atzori, F. G. Mohammadi, and K. Kochut, “Ontology summarization: Graph-based methods and beyond,” *International Journal of Semantic Computing*, vol. 13, no. 02, pp. 259–283, 2019.

[88] D. Gromann and T. Declerck, “Comparing pretrained multilingual word embeddings on an ontology alignment task,” in *LREC*, 2018.

[89] F. Scharffe, O. Zamazal, and D. Fensel, “Ontology alignment design patterns,” *Knowledge and information systems*, vol. 40, no. 1, pp. 1–28, 2014.

[90] G. Acampora, V. Loia, and A. Vitiello, “Enhancing ontology alignment through a memetic aggregation of similarity measures,” *Information Sciences*, vol. 250, pp. 1–20, 2013.

[91] A. H. Nezhadi, B. Shadgar, and A. Osareh, "Ontology alignment using machine learning techniques," *International Journal of Computer Science & Information Technology*, vol. 3, no. 2, 2011.

[92] I. Nkisi-Orji, N. Wiratunga, S. Massie, K.-Y. Hui, and R. Heaven, "Ontology alignment based on word embedding and random forest classification," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2018, pp. 557–572.

[93] F. Ardjani, D. Bouchiha, and M. Malki, "An approach for discovering and maintaining links in RDF linked data," *International Journal of Modern Education & Computer Science*, vol. 9, no. 3, 2017.

[94] M. Banane, A. Belangour, and L. El Houssine, "Storing RDF data into big data NoSQL databases," in *First International Conference on Real Time Intelligent Systems*. Springer, 2017, pp. 69–78.

[95] T. Meiser, M. Dylla, and M. Theobald, "Interactive reasoning in uncertain RDF knowledge bases," in *CIKM*, 2011, pp. 2557–2560.

[96] G. Stoilos, G. Stamou, and S. Kollias, "A string metric for ontology alignment," in *International Semantic Web conference*. Springer, 2005, pp. 624–637.

[97] M. Cheatham and P. Hitzler, "String similarity metrics for ontology alignment," in *International Semantic Web conference*. Springer, 2013, pp. 294–309.

[98] P. Jain, P. Hitzler, A. P. Sheth, K. Verma, and P. Z. Yeh, "Ontology alignment for linked open data," in *International Semantic Web conference*. Springer, 2010, pp. 402–417.

[99] J. Euzenat, "Semantic precision and recall for ontology alignment evaluation," in *IJCAI*, vol. 7, 2007, pp. 348–353.

[100] F. Scharffe and D. Fensel, "Correspondence patterns for ontology alignment," in *International Conference on Knowledge Engineering and Knowledge Management*. Springer, 2008, pp. 83–92.

[101] J. Bock and J. Hettenhausen, "Discrete particle swarm optimisation for ontology alignment," *Information Sciences*, vol. 192, pp. 152–173, 2012.

[102] K. Byrne, "Having triplets-holding cultural data as RDF," in *2008 ECDL Workshop on Information Access to Cultural Heritage*, 2008.

[103] Z. Dragisic, V. Ivanova, P. Lambrix, D. Faria, E. Jiménez-Ruiz, and C. Pesquita, "User validation in ontology alignment," in *International Semantic Web conference*. Springer, 2016, pp. 200–217.

[104] Y. Jiang, X. Wang, and H.-T. Zheng, "A semantic similarity measure based on information distance for ontology alignment," *Information Sciences*, vol. 278, pp. 76–87, 2014.

[105] D. Dzemydiene and L. Tankeleviciene, "On the development of domain ontology for distance learning course," in *EurOPT*. Vilnius Gediminas Technical University, 2008, pp. 474–479.

[106] M. Pietranik and N. T. Nguyen, "Semantic distance measure between ontology concepts attributes," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2011, pp. 210–219.

[107] E. Blanchard, M. Harzallah, H. Briand, and P. Kuntz, "A typology of ontology-based semantic measures," *EMOI-INTEROP*, vol. 160, 2005.

[108] H. Yang and J. Callan, "Learning the distance metric in a personal ontology," in *2nd International workshop on ontologies and information systems for the Semantic Web*, 2008, pp. 17–24.

[109] W. Gao, M. R. Farahani, A. Aslam, and S. Hosamani, "Distance learning techniques for ontology similarity measuring and ontology mapping," *Cluster Computing*, vol. 20, no. 2, pp. 959–968, 2017.

[110] M. Lan, J. Xu, and W. Gao, "Ontology similarity computation and ontology mapping using distance matrix learning approach," *IAENG International Journal of Computer Science*, vol. 45, no. 1, pp. 164–176, 2018.

[111] J. David and J. Euzenat, "Comparison between ontology distances (preliminary results)," in *International Semantic Web Conference*. Springer, 2008, pp. 245–260.

[112] J. Euzenat, C. Meilicke, H. Stuckenschmidt, P. Shvaiko, and C. Trojahn, "Ontology alignment evaluation initiative: Six years of experience," in *Journal on data semantics XV*. Springer, 2011, pp. 158–192.

[113] M. Ehrig, *Ontology alignment: Bridging the semantic gap*. Springer Science & Business Media, 2006, vol. 4.

[114] G. Drakopoulos, P. Mylonas, and S. Sioutas, "A case of adaptive nonlinear system identification with third order tensors in TensorFlow," in *INISTA*. IEEE, 2019, pp. 1–6.

[115] H. M. Kim, "Developing ontologies to enable knowledge management: Integrating business process and data driven approaches," in *AAAI workshop on Bringing knowledge to Business Processes*, vol. 72, 2000.

[116] M. K. Yu, J. Ma, K. Ono, F. Zheng, S. H. Fong, A. Gary, J. Chen, B. Demchak, D. Pratt, and T. Ideker, "DDOT: A swiss army knife for investigating data-driven biological ontologies," *Cell Systems*, vol. 8, no. 3, pp. 267–273, 2019.

[117] M. Hadzic and E. Chang, "Application of digital ecosystem design methodology within the health domain," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 4, pp. 779–788, 2010.